Synopsis of

Engineering structure-based RNA switches for viral detection

A THESIS TO BE SUBMITTED TO

THE UNIVERSITY OF TRANS-DISCIPLINARY HEALTH SCIENCES AND TECHNOLOGY

FOR THE AWARD OF THE DEGREE OF

DOCTOR OF PHILOSOPHY

BY

ANIRUDH NANDAKUMAR

UNDER THE GUIDANCE OF

DR. ARATI RAMESH

NATIONAL CENTRE FOR BIOLOGICAL SCIENCES (NCBS)

BANGALORE, INDIA

Table of Contents

Table of Contents ii
List of publications iii
1. Introduction 1
2. Objectives 2
3. Overview 3
3.1. Design of Toehold RNA sensors to detect SARS-CoV-2 3
3.2. Design of Toehold RNA sensors to detect viruses with diversity – Dengue 6
4. Conclusion 9
5. References 10

List of publications

Chakravarthy, Anirudh*, **Anirudh Nandakumar***, Geen George*, Shyamsundar Ranganathan, Suchitta Umashankar, Nishan Shettigar, Dasaradhi Palakodeti, Akash Gulyani, and Arati Ramesh. "Engineered RNA biosensors enable ultrasensitive SARS-CoV-2 detection in a simple color and luminescence assay." *Life Science Alliance* 4, no. 12 (2021).

Nandavaram, Aparna, **Anirudh Nandakumar**, G. M. Kashif, Annapoorni Lakshman Sagar, G. Shailaja, Arati Ramesh, and Dayananda Siddavattam. "Unusual Relationship between Iron Deprivation and Organophosphate Hydrolase Expression." *Applied and Environmental Microbiology* 89, no. 5 (2023): e01903-22.

* Co-first authors

1. Introduction

Ribonucleic acids are generally defined as a polymer of a pentose sugar – D-ribose which is linked to one of four nucleobases, and the ribose sugars are connected to one another through a phosphodiester bond. The composition was known by the end of 19th century, but it wasn't until well into the 20th century, their role as messenger-, transfer, and ribosomal- RNA was uncovered, but knowledge of the extent of their role in the cell wasn't complete until it was discovered in 1980s, that they were also capable of regulatory activities. The intra-molecular hydrogen bond network enables RNA molecules to adopt multiple conformations which differ by only a few units of energy, and this is used by genetic circuits to define and sense different states of the cell.

RNAs or parts of it, that are involved in modulating the expression of a gene present in different transcript (*trans*) or in the same transcript (*cis*), are termed regulatory RNA (whole) or regulatory RNA element (part). Inspired from numerous examples of naturally occurring regulatory RNA elements, several groups have engineered regulatory networks to cater varied purposes due to the relative ease in designing RNA elements based on thermodynamic parameters (SantaLucia, 1998; Serra & Turner, 1995; Xia et al., 1998). The versatile nature of RNA was ideal to the field of synthetic biology, a field of study that used different known regulatory elements to engineer novel functions. One of the earliest uses in synthetic biology was to report the presence of a specific transcript or small molecule in the cell. Isaacs et al., engineered an RNA that coded for GFP, in which the region containing the RBS and the start codon, base-paired with a stretch of RNA upstream to the RBS, called cis-regulatory RNA (crRNA) (Isaacs et al., 2004). A trans-acting RNA (taRNA) fragment bearing extended complementarity to the crRNA competes with the RBS for the crRNA, and promotes translation of GFP. This and other studies (Bayer & Smolke, 2005; Buskirk et al., 2004; Yen et al., 2004) that were carried out shortly after, illustrated the immense potential in designing RNA based detection platforms.

Advancements in transport and communication services, has connected even the remote places in the world with the rest, and lead to the loss and replacement of several habitats. This has led to widespread movement of people, animals, plants, and other economically valuable goods all over the planet. This has brought up the risk of infections spilling over from the non-native habitats and thus potentially starting an epidemic. Though incidences of spillover are rare, management and control of such infections requires constant surveillance in sensitive regions, and in particular for RNA viruses which are known to cross hosts and adapt to new vectors compared to other groups of pathogens.

My thesis will detail the adoption of an RNA structure-based synthetic regulatory mechanisms to develop programmable bio-sensors that are capable of reporting the presence of any target RNA of interest called Toehold RNA sensors. The Toehold RNA sensors can be designed and deploy to detect any viruses of interest even in settings with minimal infrastructure availability.

2. Objectives

The thesis will deliberate on:

- Computational workflow to design Toehold RNA biosensors from the genome sequence of a virus of interest by considering the secondary structural landscape of RNA
- Computational workflow to systematically identify regions that are conserved across genomes in viruses with known nucleotide-level diversity, to help design Toehold RNA sensors to detect all known variants of the diverse virus.

3. Overview

3.1. Design of Toehold RNA sensors to detect SARS-CoV-2

Detecting the SARS-CoV-2 virus was of paramount importance to controlling the COVID-19 pandemic, which affected people at a scale comparable to the Great flu pandemic of 1918. Public health agencies across the globe resorted to the RT-qPCR based tests, that were developed within weeks of the global outbreak of COVID-19 (Vogels et al., 2020). Though RT-qPCR based testing centres were setup in large scale, they were often swamped with samples from pouring in from all directions, since there were no reliable preliminary screening methods that eliminated the presence of infection with reasonable confidence in the initial stages. This void in preliminary screening, typically covered by immunological tests that detect circulating viral antigens or antibodies against the virus, need time to standardise and cannot be easily tailored to changing requirements. Thus, we established an RNA detection platform based on the Toehold RNA switches, that could be tailored to the target of interest based on sequence information. The Toehold RNA sensor-based detection is coupled to an isothermal amplification strategy called NASBA, that amplifies the region targeted by the Toehold sensors, thus improving their detection limit to clinically relevant levels. The Toehold RNA sensors and primers that direct amplification in NASBA, are designed through an in-house computational pipeline that only requires the genome sequence information of the virus.

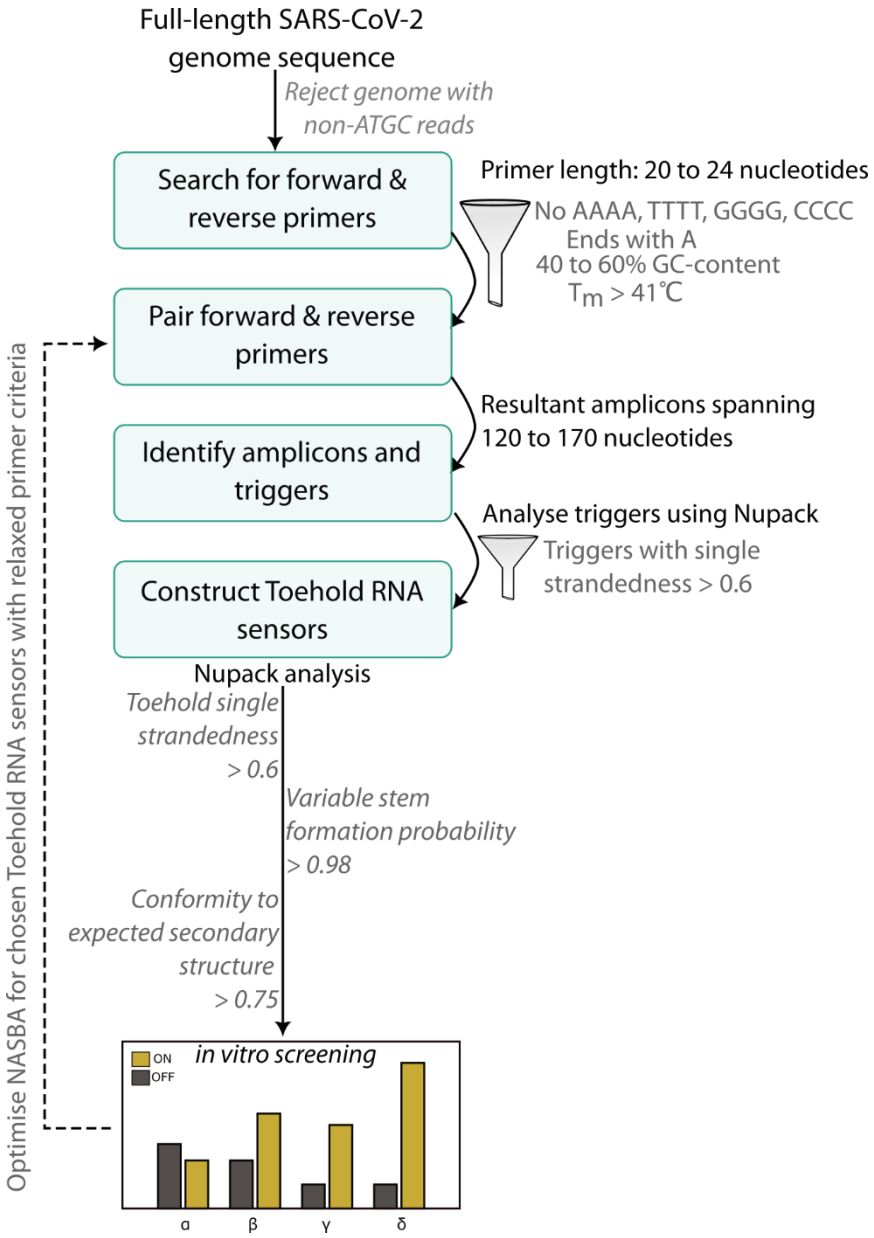
Full-length SARS-CoV-2 genome sequence
Reject genome with non-ATGC reads
Search for forward & reverse primers
Primer length: 20 to 24 nucleotides
No AAAA, TTTT, GGGG, CCCC
Ends with A
40 to 60% GC-content
Tm > 41°C
Pair forward & reverse primers
Resultant amplicons spanning 120 to 170 nucleotides
Identify amplicons and triggers
Analyse triggers using Nupack
Triggers with single strandedness > 0.6
Construct Toehold RNA sensors
Nupack analysis
Toehold single strandedness > 0.6
Variable stem formation probability > 0.98
Conformity to expected secondary structure > 0.75
Optimise NASBA for chosen Toehold RNA sensors with relaxed primer criteria
in vitro screening
ON
OFF
α
β
γ
δ

Figure 1: Outline of the computational pipeline to design Toehold RNA sensors for genomic RNA. Regions spanning 120 to 170 nucleotides in the given genome that could be amplified by a pair of primers of length 20 to 24 nucleotides were identified. Continuous sliding nucleotide windows of thirty-six nucleotides were used as a base to construct Toehold RNA sensors. Pairs of triggers and toehold RNA sensors were evaluated on the basis of four parameters that reflect on the structural properties of the Toehold RNA sensors and the trigger RNA. The Toehold RNA sensors are only designed in the regions accessible to primers to enable amplification of the region coding the target RNA by isothermal RNA amplification method – NASBA.

We implemented an *in-silico* workflow that analyses the entire genome of the viral RNA to determine targetable regions across the genome along with primers that can be

used to isothermally amplify each targetable region as illustrated in Figure 1. We shortlisted over 250 targetable regions that could be detected by over seven-hundred sensors. We tested the Toehold sensors that were coupled with *lacZ* reporter, and saw that seven of the nineteen sensors tested, showed a drastic change in colour in the presence of their cognate target RNA, and no colour change in its absence. We amplified the region targeted by one of the sensors and showed that they could detect up to 100 copies of synthetic SARS-CoV-2 RNA, efficient in detecting the presence of virus in patient samples, and the target of the sensor was conserved in ninety percent of genomes sequenced up to May 2023, and the results of the workflow have been summarised in Table 1. The workflow described here, can be employed to design Toehold RNA sensors to detect any viral RNA of interest, taking the structural aspects of RNA into consideration.

Table 1: Summary of the computational workflow to design Toehold RNA sensors to detect SARS-CoV-2 viral RNA using the genome with accession – MT012098.1.

Genome	
Length of genome	29,854
Primers	
Total possible forward, reverse primers of length 20 to 24 nucleotides	149165, 149165
Forward, reverse primers that satisfy our criteria	14305 (9.5%), 14356 (9.6%)
Amplicons	
Total possible amplicons of length 120 to 170 nucleotides	15,15,210
Amplicons identified	1,42,193 (9%)
Triggers	
Total possible triggers of 36 nucleotide length	29,819
Triggers that can code for Toehold RNA sensors	23,282
Toehold RNA sensors	
Total possible Toehold RNA sensors	1,19,276
Toehold RNA sensors designed	87,693
Triggers and Toehold RNA sensors that satisfy our criteria	
Toehold RNA sensors: 1521 (1%)	Triggers: 655 (3%)

3.2. Design of Toehold RNA sensors to detect viruses with diversity – Dengue

Viruses have been an integral part of an eco-system as seen with the association of over ~10^{13} virus like particles with an average human (Liang & Bushman, 2021; Shkoporov & Hill, 2019), and are capable of rapidly adapting to different conditions. Combined with the disruption and encroachment of different habitats, the chances of viruses spilling over and cause diseases on a global scale have been demonstrated since the Great war of the twentieth century. It has since been the interest of several health agencies to develop technologies to detect viruses that cause diseases in humans and their resources. Molecular techniques based on the detection of specific genomic regions of the viruses have been hailed as the apex detection technique by multiple health agencies. Since the viruses exhibit a high mutation rate, the key step in the nucleic acid detection kits is the identification of a target region in the virus which is highly conserved among the virus of interest and not found in other related viruses. This is typically identified though an alignment of select sequences of the entire viral genome or parts of it, and several stretches of sequence are shortlisted based on conservation, which are then screened manually to identify potential candidates for diagnosis.

The Toehold RNA sensors, as mentioned earlier is a nucleic acid based diagnostic platform that can be set up in places with low resources. The Toehold RNA sensors so far have been implemented to detect different ssRNA viruses such Zika virus (Pardee et al., 2016), Norovirus (Ma et al., 2018), and SARS-CoV-2 virus (Amalfitano et al., 2021; Chakravarthy et al., 2021; Park & Lee, 2021); none of the approaches reported systematically account for the diversity exhibited by RNA viruses. Figure 2 shows the percentage of sequenced genomes covered by all possible thirty-six nucleotide trigger regions generated from the reference genome(s) of Zika-, Noro-, Dengue-, and MERS-CoV viruses. This shows that the diversity observed in the viruses needs to be incorporated in the design of Toehold RNA sensors especially for cases like Dengue and Norovirus.

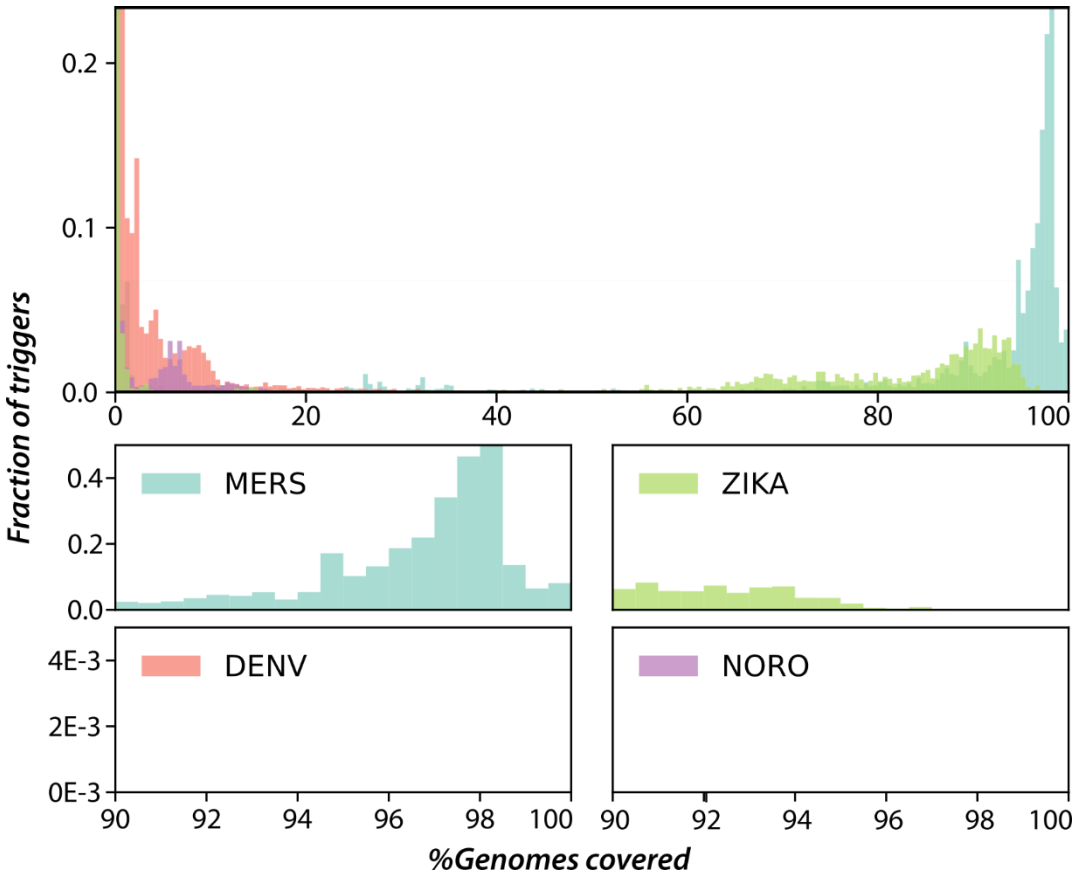

Figure 2: Extent of genomes covered by triggers across multiple viruses. The extent of fully-sequenced genomes covered by all possible thirty-six nucleotide (triggers) sequences generated from the reference genome(s) of MERS, Zika, Dengue, and Noro virus. Top panel – showing the coverage of all four viruses. Bottom panels – showing the coverage in each of the viruses spanning from 90 to 100% of genomes belonging to each group.

The lack of conserved Toehold RNA target regions poses an interesting challenge, but it is known that nucleic acid detection tests such as RT-qPCR often tolerate mismatches in sequence complementarity between the primer and target (Kwok et al., 1990; Lefever et al., 2013), albeit with a concomitant reduction in the signal observed. In addition, *Ma et al.,* (Ma et al., 2018) showed that the Toehold RNA sensors could tolerate up to 4 mismatches, but the sequence with 4 mismatches (norovirus GII.6 against Toehold sensor S2) showed only half the response compared to the un-mutated sequence (norovirus GII.4 Sydney against Toehold sensor S2). Though it was not shown systematically, we chose to consider stretches of genomes that were different by up to 3 mismatches of the trigger being considered, and these stretches were termed *similar regions*.

To identify conserved regions that could be targeted by Toehold RNA sensors, we calculated the *prevalence* of a trigger which was the sum of the number of genomes in which a specific trigger was present, and the number of genomes in which the regions similar to the trigger were present (Figure 3).

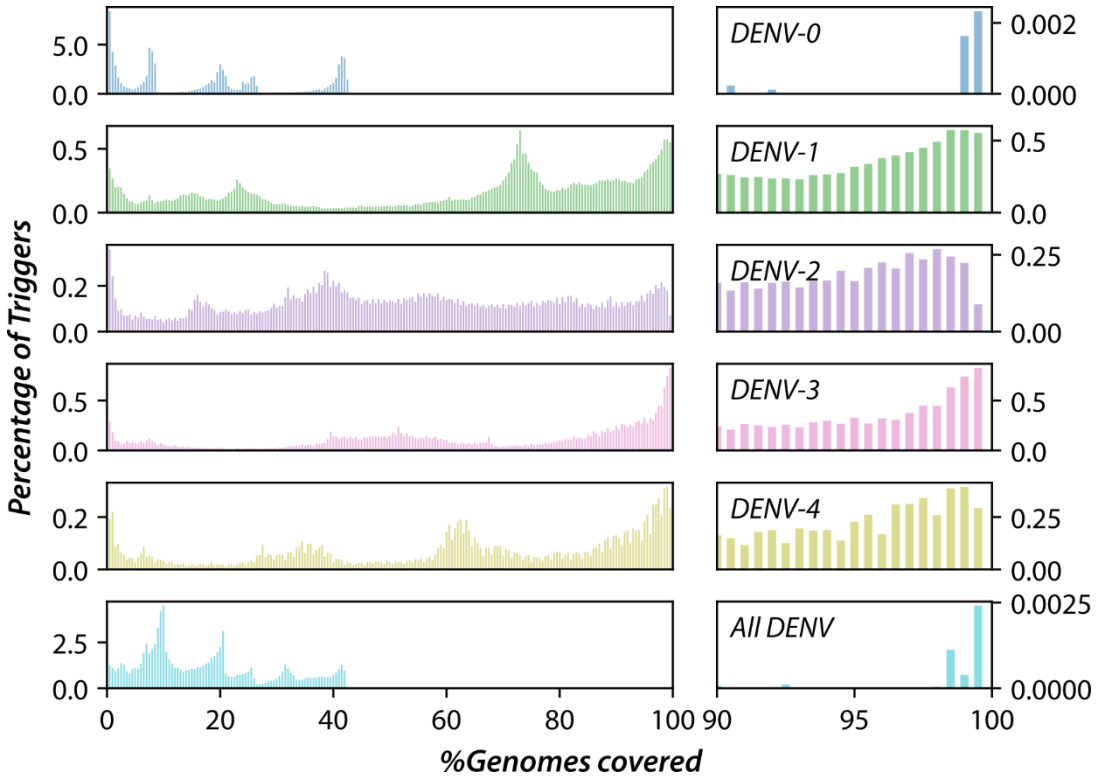

Figure 3: Prevalence of triggers across Dengue genomes. The prevalence of each trigger generated from all 6712 fully-sequenced genomes of Dengue virus. The panels on the right highlight the percentage of triggers with prevalence in more than 90% of genomes of each group. Note for panels on the left: Triggers with prevalence in less than 1% of the genomes have not been included in the analysis.

Results of the computational analysis to identify pan-Dengue Toehold RNA sensors is summarised in Table 2. We chose 21 Toehold RNA sensors that detected 13 different pan-Dengue triggers to be tested *in vitro*, and found eight Toehold RNA sensors responded specifically to their cognate triggers with high signal to noise ratio. We confirmed our assumptions by testing three of these sensors with majority of the variations found in the Dengue genome, and found that they could detect more than 99% of all the fully sequenced Dengue genomes.

Table 2: Summary of the computational pipeline to identify pan-Dengue Toehold RNA sensors.

Total genomes analysed	6712
Total possible triggers	7.03×10^{7}
Total unique triggers	2.57×10^{6} (3.7%)
Total pan-Dengue triggers	61
Total Toehold RNA sensors targeting pan-Dengue triggers	219
Pan-Dengue triggers that met the selection criteria	29 (47.5%)
Toehold RNA sensors that met the selection criteria	67 (30.6%)
Pan-Dengue triggers that were tested in vitro	14 (22.9%)
Toehold RNA sensors that were tested	21 (9.6%)

4. Conclusion

Together, with the computational workflows proposed in this thesis, it is possible to design and develop nucleic acid detection platforms which can also work in low resource

settings, and they both serve different purposes. Reports of RNA viruses crossing multiple hosts has increased in recent times; surveillance systems and diagnostic systems play a crucial role in preventing and managing epidemics, and the approaches to design point of care nucleic acid diagnostics described in this thesis will help in designing better diagnostics to detect emerging and existing viruses.

5. References

1. Amalfitano, E., Karlikow, M., Norouzi, M., Jaenes, K., Cicek, S., Masum, F., Sadat Mousavi, P., Guo, Y., Tang, L., Sydor, A., Ma, D., Pearson, J. D., Trcka, D., Pinette, M., Ambagala, A., Babiuk, S., Pickering, B., Wrana, J., Bremner, R., … Pardee, K. (2021). A glucose meter interface for point-of-care gene circuit-based diagnostics. *Nature Communications*, *12*(1), 724. https://doi.org/10.1038/s41467-020-20639-6

2. Bayer, T. S., & Smolke, C. D. (2005). Programmable ligand-controlled riboregulators of eukaryotic gene expression. *Nature Biotechnology*, *23*(3), 337–343. https://doi.org/10.1038/nbt1069

3. Buskirk, A. R., Landrigan, A., & Liu, D. R. (2004). Engineering a Ligand-Dependent RNA Transcriptional Activator. *Chemistry & Biology*, *11*(8), 1157–1163. https://doi.org/10.1016/j.chembiol.2004.05.017

4. Chakravarthy, A., Nandakumar, A., George, G., Ranganathan, S., Umashankar, S., Shettigar, N., Palakodeti, D., Gulyani, A., & Ramesh, A. (2021). Engineered RNA biosensors enable ultrasensitive SARS-CoV-2 detection in a simple color and luminescence assay. *Life Science Alliance*, *4*(12). https://doi.org/10.26508/lsa.202101213

5. Isaacs, F. J., Dwyer, D. J., Ding, C., Pervouchine, D. D., Cantor, C. R., & Collins, J. J. (2004). Engineered riboregulators enable post-transcriptional control of gene expression. *Nature Biotechnology*, *22*(7), 841–847. https://doi.org/10.1038/nbt986

6. Kwok, S., Kellogg, D. E., McKinney, N., Spasic, D., Goda, L., Levenson, C., & Sninsky, J. J. (1990). Effects of primer-template mismatches on the polymerase chain

reaction: Human immunodeficiency virus type 1 model studies. *Nucleic Acids Research*, *18*(4), 999–1005. https://doi.org/10.1093/nar/18.4.999

7. Lefever, S., Pattyn, F., Hellemans, J., & Vandesompele, J. (2013). Single-Nucleotide Polymorphisms and Other Mismatches Reduce Performance of Quantitative PCR Assays. *Clinical Chemistry*, *59*(10), 1470–1480. https://doi.org/10.1373/clinchem.2013.203653

8. Liang, G., & Bushman, F. D. (2021). The human virome: Assembly, composition and host interactions. *Nature Reviews Microbiology*, *19*(8), 514–527. https://doi.org/10.1038/s41579-021-00536-5

9. Ma, D., Shen, L., Wu, K., Diehnelt, C. W., & Green, A. A. (2018). Low-cost detection of norovirus using paper-based cell-free systems and synbody-based viral enrichment. *Synthetic Biology*, *3*(1), ysy018. https://doi.org/10.1093/synbio/ysy018

10. Pardee, K., Green, A. A., Takahashi, M. K., Braff, D., Lambert, G., Lee, J. W., Ferrante, T., Ma, D., Donghia, N., Fan, M., Daringer, N. M., Bosch, I., Dudley, D. M., O'Connor, D. H., Gehrke, L., & Collins, J. J. (2016). Rapid, Low-Cost Detection of Zika Virus Using Programmable Biomolecular Components. *Cell*, *165*(5), 1255–1266. https://doi.org/10.1016/j.cell.2016.04.059

11. Park, S., & Lee, J. W. (2021). Detection of Coronaviruses Using RNA Toehold Switch Sensors. *International Journal of Molecular Sciences*, *22*(4), Article 4. https://doi.org/10.3390/ijms22041772

12. SantaLucia, J. (1998). A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proceedings of the National Academy of Sciences*, *95*(4), 1460–1465. https://doi.org/10.1073/pnas.95.4.1460

13. Serra, M. J., & Turner, D. H. (1995). [11] Predicting thermodynamic properties of RNA. In *Methods in Enzymology* (Vol. 259, pp. 242–261). Academic Press. https://doi.org/10.1016/0076-6879(95)59047-1

14. Shkoporov, A. N., & Hill, C. (2019). Bacteriophages of the Human Gut: The "Known Unknown" of the Microbiome. *Cell Host & Microbe*, *25*(2), 195–209. https://doi.org/10.1016/j.chom.2019.01.017

15. Vogels, C. B. F., Brito, A. F., Wyllie, A. L., Fauver, J. R., Ott, I. M., Kalinich, C. C., Petrone, M. E., Casanovas-Massana, A., Catherine Muenker, M., Moore, A. J., Klein, J., Lu, P., Lu-Culligan, A., Jiang, X., Kim, D. J., Kudo, E., Mao, T., Moriyama, M., Oh, J. E., ... Grubaugh, N. D. (2020). Analytical sensitivity and efficiency comparisons of SARS-CoV-2 RT–qPCR primer–probe sets. *Nature Microbiology*, *5*(10), 1299–1305. https://doi.org/10.1038/s41564-020-0761-6

16. Xia, T., SantaLucia, J. Jr., Burkard, M. E., Kierzek, R., Schroeder, S. J., Jiao, X., Cox, C., & Turner, D. H. (1998). Thermodynamic Parameters for an Expanded Nearest-Neighbor Model for Formation of RNA Duplexes with Watson–Crick Base Pairs. *Biochemistry*, *37*(42), 14719–14735. https://doi.org/10.1021/bi9809425

17. Yen, L., Svendsen, J., Lee, J.-S., Gray, J. T., Magnier, M., Baba, T., D'Amato, R. J., & Mulligan, R. C. (2004). Exogenous control of mammalian gene expression through modulation of RNA self-cleavage. *Nature*, *431*(7007), 471–476. https://doi.org/10.1038/nature02844